



Memoria del XLIX Taller de Actualización Bioquímica, Facultad de Medicina; UNAM

Técnicas para el estudio del microbioma.

Techniques for the study of the microbiome.

Valdez-Palomares, Jessica Fernanda^{1*}.

1. Universidad Nacional Autónoma de México. Posgrado en Ciencias Biológicas. Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán. Laboratorio de Fisiología de la Nutrición e Instituto Nacional de Medicina Genómica. Unidad de Vinculación Científica de la Facultad de Medicina UNAM-INMEGEN.

*Correspondencia: INMEGEN, Periferico Sur 4809, Arenal Tepepan, Tlalpan, C.P. 14610, CDMX, México.
Tel +52(55)53501900 ext 1200, palomares.fernanda@gmail.com

Resumen

La microbiota intestinal es un tema de estudio novedoso, cuyo número de publicaciones ha crecido rápidamente, dada la amplia cantidad de observaciones que señalan la relación entre las bacterias intestinales y el estado de salud o enfermedad en humanos. La utilización de muestras fecales, es una alternativa metodológica de gran utilidad para estudiar las comunidades microbianas del intestino por su fácil obtención de carácter no invasivo. Sin embargo, el uso de estas muestras comprende distintas consideraciones metodológicas que abarcan la obtención y condiciones de almacenamiento y procesamiento de las muestras como importantes variables a considerar, para poder preservar las características originales de la muestra y su representatividad. Así como la estrategia de secuenciación usando tecnologías de nueva generación y el cómo podemos valernos de recursos bioinformáticos para ayudarnos a conocer la diversidad microbiana de las secuencias derivadas de estas muestras y la presencia de taxones con abundancia diferencial. Por lo tanto, el conocimiento de los retos asociados a la obtención, almacenamiento y procesamiento de muestras fecales y la disponibilidad de técnicas de secuenciación y herramientas bioinformáticas para los análisis de diversidad y taxonomía, puede ayudarnos a mejorar el diseño experimental al estudiar a la microbiota intestinal.

Palabras claves: microbiota intestinal, heces, 16S, conservación, protocolo, secuenciación, amplicón, diversidad.

Abstract

The intestinal microbiota has garnered increasing interest in recent years, given the many observations that point to the relationship between intestinal bacteria and health or disease in humans. The use of fecal samples is a helpful alternative to study the microbial communities of the intestine due to their non-invasive nature. However, using these samples involves different methodological considerations that include the acquisition, storage, and processing conditions of the samples as major variables to be taken into account to preserve its representativeness. As well as the sequencing strategy using new generation technologies and how we can use bioinformatics resources to help us to determine the microbial diversity of the sequences derived from these samples and the presence of differentially abundant taxa. Therefore, knowledge of the challenges associated with collecting, storing, and processing fecal samples and the availability of sequencing techniques and bioinformatics tools for diversity and taxonomic analyses can help us to improve the experimental design when studying gut microbiota.

Keywords: intestinal microbiota, feces, 16S, storage, protocol, sequencing, amplicon, diversity.

Introducción

Los análisis taxonómicos de muestras fecales son ampliamente utilizados para obtener una aproximación a la ecología microbiana del tracto gastrointestinal.

La secuenciación de un fragmento corto del gen 16SrRNA en las regiones hipervariables V₃-V₄ o más recientemente, solo la región V₄, permite obtener, una robusta estimación del perfil taxonómico y la diversidad bacteriana hasta el nivel de género (1). Las muestras fecales han sido ampliamente utilizadas

en diversos estudios y ensayos clínicos al ser fáciles de coleccionar por su carácter de obtención poco invasiva.

A continuación, se presentan alternativas y consideraciones metodológicas para el estudio de la microbiota intestinal que comprenden, la obtención y almacenamiento de materia fecal, purificación de DNA, amplificación y secuenciación del gen 16SrRNA en las regiones hipervariables V₃-V₄ y una breve aproximación al análisis bioinformático de las secuencias obtenidas.

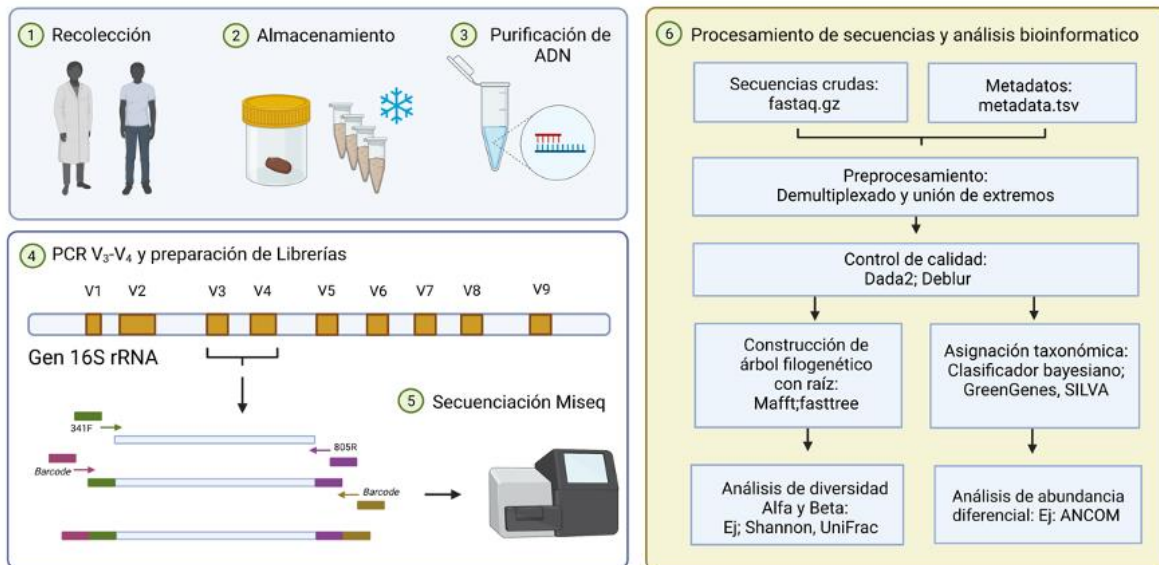


Figura 1. Resumen de diagrama de flujo para análisis de microbiota intestinal proveniente de muestras fecales. 1) Apoyo por parte del personal de salud para la recolección de muestras fecales. 2) Almacenamiento de muestras fecales por congelación inmediata. 3) Purificación de ADN. 4) Amplificación por técnica de PCR de las regiones hipervariables V₃-V₄ del gen y preparación de librerías para secuenciación 16S rRNA. Se puede observar el gen 16S rRNA y sus regiones hipervariables. 5) Secuenciación en Miseq. 6) Procesamiento de secuencias crudas y análisis bioinformático.

Obtención de muestras fecales

En ensayos clínicos, es ideal que el personal de salud instruya a los pacientes con un protocolo escrito, acompañado de elementos de apoyo visual para el adecuado manejo y almacenamiento de heces (Fig. 1.1), que serán entregadas de acuerdo a las necesidades del protocolo a realizar (2). Los protocolos de recolección y conservación de muestras afectan tanto a la calidad como a la precisión de los datos metagenómicos (3). El tamaño del efecto de estos pasos, en algunas circunstancias, puede ser mayor que el de las variables biológicas de interés y pueden ser importantes factores de confusión en los meta-análisis de conjuntos de datos de diferentes estudios (Fig. 1.2). Los métodos de

colecta y almacenamiento que han sido validados para un tipo de muestra, no pueden asumirse como óptimos para todos los casos. Por ello, a menudo es necesario optimizarlos de acuerdo a los objetivos clave, que son; recoger suficiente biomasa microbiana para la secuenciación y minimizar la contaminación de las muestras (4). La congelación inmediata, sin conservante de heces, se ha utilizado ampliamente como estándar de oro para los análisis del microbioma intestinal, ya que este método preserva la composición microbiana de forma similar a una muestra fresca y también evita la posible influencia del conservante añadido (5). Por otro lado, se han reportado otros sistemas alternativos de colecta de muestras que no requieren congelación, incluidos OMNIgene GUT, etanol al 95%, RNAlater y tarjetas FTA, que también son informativos para el

análisis del microbioma intestinal (5). De estos el etanol al 95% tiene una validez relativamente baja para los índices de alfa diversidad (6).

Extracción de DNA

Actualmente podemos encontrar en el mercado un sinnúmero de kits comerciales para la purificación de ADN para estudios de microbiota. Es importante mencionar que la metodología de extracción del ADN puede afectar a la composición de los datos de la secuencia descendente (7) (Fig. 1.3). Los métodos de extracción de ADN que incluyen la lisis mecánica (o el batido de microesferas) suelen considerarse superiores a los que se basan en la lisis química (8). Sin embargo, varían en cuanto a su eficacia (9). La contaminación puede ocurrir durante las etapas de procesamiento de la muestra y los kits o reactivos de laboratorio pueden contener cantidades variables de contaminantes microbianos (10). Para la purificación de ADN proveniente de heces frescas o congeladas, pueden usarse kits que utilizan columnas que integran membranas de sílice. El procedimiento no requiere extracción con fenol cloroformo, y supone una manipulación mínima que permite reducir la contaminación cruzada entre muestras y facilitar el manejo de muestras potencialmente infecciosas. Estos procedimientos están optimizados para su uso con cantidades que van de 180 a 220 mg de heces, pero también pueden utilizarse con volumen mayor, cuando el ADN diana no está distribuido de forma homogénea en las heces y/o se encuentra en una baja concentración (11).

Las células bacterianas se lisan eficazmente incubando el homogeneizado de heces con amortiguador de lisis a 70°C, pero para la detección de células difíciles de lisar esta temperatura puede aumentarse a 95°C. El método de extracción debe ser eficaz para diversos taxones microbianos; de lo contrario, los resultados de la secuenciación pueden estar dominados por el ADN procedente únicamente de microbios fáciles de lisar (7).

Las muestras de heces suelen contener muchos compuestos, que pueden degradar el ADN e inhibir las reacciones enzimáticas posteriores como la Reacción en Cadena de la Polimerasa (en inglés "PCR"), entre los inhibidores más comunes en las heces se encuentran los ácidos biliares, uratos y los polisacáridos complejos (12). Existen kits con tabletas o soluciones amortiguadoras para remover sustancias inhibitorias, después de los pasos de lisis, estos pueden eliminarse mediante centrifugación, el ADN se purifica en columnas en pasos posteriores.

El procedimiento de purificación del ADN implica (11):

1. *Digestión de las proteínas.* Con proteinasa K en condiciones desnaturalizantes.
2. *La unión del ADN a una membrana de sílice.* Se ajustan las condiciones con un amortiguador para permitir la unión del ADN a la membrana, y la muestra se carga en la columna de centrifugado. Donde el ADN se absorbe en la membrana de sílice durante un paso de centrifugación.
3. *Lavado de las impurezas.* Las concentraciones de sal y las condiciones de pH optimizadas en el lisado evitan que los restos de proteínas digeridas y otras impurezas, que pueden inhibir la PCR queden retenidas en la membrana. El ADN unido a la membrana se lava mediante centrifugación con tampones de lavado.
4. *Elución del ADN.* El ADN se eluye de la columna de centrifugación con un tampón de baja salinidad a temperatura ambiente, libre de proteínas, nucleasas impurezas o inhibidores de la PCR.

Determinación de la concentración, el rendimiento y la pureza

El ADN purificado puede cuantificarse por espectrofotometría UV visible de espectro completo. El rendimiento del ADN se determina a partir de la concentración de ADN en el eluido, medida por la absorbancia a longitud de onda de 260 nm (13). La pureza se determina calculando la relación entre la absorbancia a 230nm, 260nm y 280nm. Para evaluar con precisión la calidad de la muestra, deben analizarse las relaciones 260/280 y 260/230 en combinación con la calidad espectral general. Los ácidos nucleicos puros suelen arrojar una relación 260/280 de ~1.8 y de ~2.0 para el ADN y el ARN, respectivamente. Esta relación depende del pH y de la fuerza iónica del amortiguador utilizado para realizar las mediciones del blanco y de la muestra. Las soluciones ácidas subestiman la relación en 0.2-0.3, mientras que una solución básica la sobreestima en 0.2-0.3. Las relaciones de pureza significativamente diferentes pueden indicar la presencia de proteínas, fenol u otros contaminantes que absorben fuertemente a 280 nm. La relación de pureza 260/230 es una segunda medida de la pureza del ADN, con valores para un ácido nucleico puro que suelen estar en el rango de 1.8-2.2. Los cocientes de pureza que son significativamente inferiores a los valores esperados, pueden indicar que la técnica de

aislamiento utilizada puede requerir una mayor optimización (14).

Tanto el ADN como el ARN se miden con un espectrofotómetro. Para medir sólo el ADN, se debe utilizar un fluorómetro. La integridad y longitud del ADN se determina, haciendo una electroforesis en un gel de agarosa al 2%. El ADN puede teñirse con un agente intercalante para su visualización.

Amplificación de las regiones hipervariables V3-V4 del gen 16S rRNA y purificación de amplicones

La PCR se ha convertido en una importante herramienta para la identificación de microorganismos previamente desconocidos y los análisis de diversidad microbiana ambiental (15). La PCR es utilizada para amplificar las regiones hipervariables V₃-V₄ del gen 16S rRNA a partir del ADN purificado de muestras fecales. Para esto, se recomienda utilizar una polimerasa de alta fidelidad y los primers Bakt_341F y Bakt_805R (16) (Tabla 1), generando fragmentos de ~500 pb a partir de 1-200 ng de DNA de cada muestra (Fig. 1.4).

Tabla 1. Cebadores Bakt_341F y Bakt_805R compatibles con Illumina.

Forward	5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG CTACGGGNGGCWGCAG 3'
Reverse	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG GACTACHVGGGTATCTAATCC 3'

Las polimerasas de alta fidelidad son eficaces hasta tamaños de 15 kb de ADN genómico. Podemos encontrar en el mercado mezclas maestras (*Master Mix*) que contienen anticuerpos anti-Taq ADN polimerasa, Mg⁺⁺, desoxirribonucleótidos trifosfatos, ADN polimerasa Taq recombinante y de *Pyrococcus* GB-D termoestable. El anticuerpo anti-Taq ADN polimerasa inhibe la actividad de la polimerasa de forma reversible (17), proporcionando un "hot start" automatizado y permitiendo una configuración a temperatura ambiente. El hot start mediado por anticuerpos mejora la especificidad y el rendimiento de la PCR. La actividad de la polimerasa se restablece tras un paso de desnaturalización en el ciclo de PCR a 94°C. La polimerasa GB-D de la especie *Pyrococcus* posee una actividad de corrección de exonucleasas de 3' a 5' (18). La mezcla de la enzima correctora con la Taq ADN polimerasa aumenta la fidelidad ~6 veces sobre la de la Taq ADN polimerasa sola.

Los amplicones resultantes pueden purificarse por centrifugación mediante membranas de sílice, o perlas magnéticas.

Secuenciación de las regiones hipervariables V₃-V₄ del gen 16S rRNA

Los amplicones pueden secuenciarse utilizando metodologías de alto rendimiento, por ejemplo, la plataforma Illumina, para la preparación de librerías de secuenciación metagenómica de 16S (19). Según el protocolo de preparación de librerías, se amplifican las regiones hipervariables V₃-V₄ en un ciclo de PCR limitado y se agregan los adaptadores de secuenciación y códigos de barra (*Barcodes*). Las bibliotecas se agrupan para su secuenciación utilizando los índices de NexteraXT. Usando lecturas pareadas de 300 pb, los extremos de cada par de lecturas se sobreponen, para generar lecturas completas de alta calidad, en una corrida de aproximadamente 56 horas. El resultado de corrida en MiSeq es de aproximadamente 44-50 millones de lecturas de extremos pareados o *paired-end* (Fig. 1.5).

Procesamiento de secuencias y control de calidad

Para el análisis bioinformático a partir de las secuencias obtenidas del 16S rRNA, puede utilizarse la plataforma Quantitative Insights Into Microbial Ecology 2 (QIIME2) en su versión más actual (20) (Fig. 1.6).

Primero se genera un archivo que contenga los metadatos asociados a las muestras o *Metadata.tsv*, para importar las secuencias crudas pareadas. Los metadatos son datos contextuales sobre los datos experimentales. Son el quién, el qué, cuándo, dónde y por qué de estos datos. Los metadatos de la muestra son la información sobre la muestra primaria: cuándo se recogió (por ejemplo, fecha y hora), de dónde se recogió (por ejemplo, latitud, longitud, elevación/profundidad, nombre del sitio, país, etc.), qué tipo de muestra era (por ejemplo, suelo, agua de mar, heces) y las propiedades del entorno durante la colecta (por ejemplo, temperatura, salinidad, pH) o la condición experimental (por ejemplo, experimental o de control, estado de enfermedad) de la que se tomó la muestra. Los metadatos de preparación incluyen información sobre los métodos de preparación y/o procesamiento de la muestra primaria. Para la secuenciación de ADN de una muestra del microbioma, los metadatos de preparación incluyen el tipo de método de extracción de ADN, y las condiciones utilizadas para la secuenciación (por

ejemplo, cebadores, kits de bibliotecas, instrumentación y parámetros de secuenciación) (21).

Las hipótesis son impulsadas por los metadatos y dependen de ellos. En el contexto de las muestras de microbiota, las variables descriptivas permiten hacer comparaciones (por ejemplo, entre grupos, con otras covariables, etc.). Los metadatos también son fundamentales para compartir y reutilizar los datos científicos. Los principios de disposición de datos están recogidos en los Principios de Datos FAIR (22), que establecen que los datos deben ser localizables, accesibles, interoperables y

reutilizables. Si se tienen secuencias multiplexadas *paired-end* con *barcodes*, se pueden generar dos archivos que contengan las lecturas forward y reverse, bajo el mismo nombre que se les otorgó en el archivo metadata, el archivo metadata.tsv debe contener una columna con los *barcodes* asociados, o dos columnas si son índices duales (Fig. 2). El orden de los registros en los archivos *fastq*, define la asociación entre las lecturas de la secuencia forward y reverse, por lo que debe conservarse un orden correcto. No es necesario que los códigos de barras del archivo de metadatos estén en un orden específico.

	A	B	C	D	E	F
1	sample-id	barcode-sequence	body-site	year	month	day
2	#q2:types	categorical	categorical	numeric	numeric	numeric
3	RAT1	AGCTGACTAGTC	COLON	2021	12	28
4	RAT2	ACACACTATGGC	COLON	2022	1	20
5	RAT3	ACTACGTGTGGT	COLON	2022	1	17
6	RAT4	AGTGCGATGCGT	COLON	2022	1	17
7	RAT5	ACGATGCGACCA	JEJUNUM	2022	1	20
8	RAT6	AGCTATCCACGA	JEJUNUM	2022	1	17
9	RAT7	ATGCAGCTCAGT	JEJUNUM	2022	1	17
10	RAT8	CACGTGACATGT	JEJUNUM	2022	1	14

Figura 2. Ejemplo de archivo de metadatos para importar secuencias en QIIME2. La primera fila debe contener el nombre de la variable, siendo la primera celda nombrada como "sample-id". La segunda fila indica el tipo de variable (categórica o numérica), la primera celda debe ser nombrada como "#q2:types".

Para importar los datos a QIIME2 puede usarse el comando `qiime tools import`. Una vez importados los archivos se comienza la demultiplexación de secuencias. Para ello es necesario el archivo de metadatos de la muestra, y debe indicar qué columna de ese archivo contiene los códigos de barras por muestra.

Tras la demultiplexación, podemos generar y visualizar un resumen de cuántas secuencias se obtuvieron por muestra y la calidad de la secuencia en cada posición.

Se puede escoger entre distintos métodos de control de calidad, como DADA2 y Deblur. Dado que las ASVs resultantes de DADA2 se crean agrupando secuencias únicas se denominan variantes de secuencia del amplicón o ASVs. DADA2 permite detectar y corregir los datos de la secuencia de amplicones de Illumina, este proceso de control de calidad filtra además cualquier lectura phiX (comúnmente presente en los datos de secuencias de genes marcadores de Illumina) que se identifique en los datos de secuenciación, y filtra las secuencias quiméricas (23,24).

Pruebas de diversidad

QIIME provee la implementación de varias métricas de diversidad filogenética. Estas métricas requieren la generación de un árbol filogenético enraizado. El cual puede generarse usando el comando `align-to-tree-mafft-fasttree` del plugin `q2-phylogeny`.

Para realizar análisis de diversidad, podemos usar el comando `q2-diversity` plugin y el método `core-metrics-phylogenetic` que calcula varias métricas de diversidad alfa y beta, y genera gráficos de análisis de coordenadas principales (PCoA). Un parámetro muy importante que debe proporcionarse es la profundidad de muestreo uniforme, es decir, de rarefacción. Porque la mayoría de las métricas de diversidad son sensibles a diferentes profundidades de muestreo en diferentes muestras. La rarefacción es un método propuesto por primera vez por Sanders (25), que ajusta las diferencias de tamaño de las bibliotecas entre las muestras. Basándose en estas submuestras de igual tamaño, se pueden calcular métricas de diversidad que pueden contrastar los

ecosistemas de forma "justa", independientemente de las diferencias en los tamaños de las muestras (26).

La elección de este valor es complicada, una opción es elegir un valor lo más alto posible para retener más secuencias por muestra y excluir el menor número de muestras posible. Se ha reportado que las estimaciones de muchos índices de diversidad alfa subestiman la verdadera diversidad del entorno (27). Intentar abordar este problema utilizando la rarefacción en ocasiones puede inducir más sesgo, porque los entornos pueden ser idénticos con respecto a una métrica de diversidad alfa, pero las diferentes estructuras de abundancia inducirán diferentes sesgos cuando se enrarezcan. Por ejemplo, en dos entornos con diferentes estructuras de abundancia pero con igual riqueza; la rarefacción puede dar la falsa impresión de una riqueza desigual (28). Tanto la riqueza de la muestra como la riqueza enrarecida están impulsadas por artefactos del experimento como el tamaño de la biblioteca, y no puramente por la estructura de la comunidad microbiana. Recientemente se argumenta que, para sacar conclusiones significativas sobre toda la comunidad microbiana, es necesario ajustar el muestreo no exhaustivo utilizando modelos de error de medición para ajustar la incertidumbre en la estimación de la diversidad alfa. Por lo que se ha motivado a los ecologistas microbianos a utilizar estimaciones de la diversidad alfa que tengan en cuenta las especies no observadas, y a utilizar la varianza de las estimaciones en los modelos de error de medición para comparar la diversidad entre ecosistemas (29)

Podemos analizar la composición de las muestras en el contexto de los metadatos categóricos mediante PERMANOVA. (análisis de varianza multivariante permutado) (30), que es una partición geométrica de la variación multivariante en el espacio de una medida de disimilitud elegida según un diseño ANOVA dado, con valores p obtenidos mediante técnicas de permutación libres de distribución. El método es semiparamétrico, motivado por el deseo de realizar una partición clásica, como en el ANOVA (permitiendo así las pruebas y la estimación de los tamaños de los efectos principales, los términos de interacción, las estructuras jerárquicas, los componentes aleatorios en los modelos mixtos, etc.) conservando al mismo tiempo importantes propiedades estadísticas robustas de los métodos multivariantes no paramétricos basados en rangos, como el análisis de similitudes (ANOSIM) (31), como, la flexibilidad para basar el análisis en una medida de disimilitud (como Bray-Curtis, Jaccard, etc.) y las inferencias sin distribución logradas por

permutaciones, sin la suposición de normalidad multivariante. Puede haber más variables de respuesta que unidades de muestreo, los datos pueden ser severamente no normales, cero-inflados, ordinales o cualitativos (30). Prueba si las distancias entre las muestras dentro de un grupo, son más similares entre sí que con las muestras de otros grupos, esta prueba puede ser lenta de ejecutar, especialmente si es pareada, al basarse en pruebas de permutación.

Análisis taxonómico

Podemos explorar la composición taxonómica de las muestras en función de los metadatos. QIIME2 permite asignar la taxonomía a las secuencias utilizando un clasificador Naive Bayes y el plugin `q2-feature-classifier`. Los clasificadores taxonómicos funcionan mejor cuando se entrenan en función de la preparación de la muestra y los parámetros de secuenciación específicos, incluidos los cebadores que se utilizaron para la amplificación y la longitud de las lecturas de la secuencia, estos clasificadores también pueden entrenarse con distintas bases de datos como SILVA y Green Genes.

La determinación de bacterias diferencialmente abundantes entre dos o más ambientes, representa un reto que ha recibido gran interés. Los datos observados del microbioma (tabla OTU/SV) son abundancias relativas con un exceso de ceros. Dado que las abundancias relativas suman una constante, estos datos son composicionales. El estudio del microbioma se trata de un ámbito de investigación que crece muy rápidamente y en el que se desarrollan regularmente nuevos métodos estadísticos. Especialmente en el caso de los métodos de análisis de abundancia diferencial. Existen varios métodos en la literatura y cada uno de ellos tiene sus propios puntos fuertes y débiles, y debe considerarse que no todos ellos están diseñados para probar hipótesis estadísticas relativas al mismo parámetro. Algunos métodos están diseñados para probar hipótesis sobre la abundancia relativa, mientras que otros están diseñados para probar hipótesis sobre la abundancia absoluta y su elección afecta la tasa de descubrimientos falsos (FDR). Un problema relacionado es que se usan los términos "abundancia relativa" y "abundancia absoluta en una unidad de volumen" indistintamente. Por ello, es importante precisar la terminología.

El análisis ANCOM (Análisis de Composiciones de Microbiomas) puede aplicarse para identificar características que son diferencialmente abundantes (es decir, que están presentes en diferentes

abundancias) entre los grupos de muestras (32). ANCOM asume que pocas características (menos del 25%) cambian entre los grupos. Si esperamos que haya más características que cambian entre los grupos, no debemos utilizar ANCOM, ya que será más propenso a los errores de tipo I y de tipo II. ANCOM se basa en las frecuencias de las características por muestra, pero no puede tolerar frecuencias de cero, por lo tanto, requiere un método de imputación. A menudo nos interesa realizar una prueba de abundancia diferencial a un nivel taxonómico específico (32). Para ello, podemos contraer las características en el nivel taxonómico de interés al ejecutar pruebas de abundancia diferencial.

Recientemente, se propuso una metodología de ANCOM con corrección de Sesgo (ANCOM-BC), que estima las fracciones de muestreo desconocidas y corrige el sesgo inducido por sus diferencias entre muestras. Los datos de abundancia absoluta se modelan utilizando un marco de regresión lineal (33). Se han realizado estudios de simulación para comparar el FDR de ANCOM como ANCOM-BC, únicamente fallan no controlando el FDR cuando el tamaño de las muestras es muy pequeño, como <10 (33). Otros métodos como Clasificación Diferencial (DR), DESeq2, edgeR y ALDEx2 tienden a inflar el FDR para todos los tamaños de muestra y su FDR empeora con el aumento del tamaño de la muestra (33). Esto se debe a que, bajo la hipótesis nula, cada uno de estos métodos está sesgado lejos de cero. Este sesgo aumenta con el tamaño de la muestra. Por lo tanto, el FDR aumenta con el tamaño de la muestra.

Referencias

- Liu, P.-Y., Wu, W.-K., Chen, C.-C., Panyod, S., Sheen, L.-Y., and Wu, M.-S. (2020) Evaluation of Compatibility of 16S rRNA V3V4 and V4 Amplicon Libraries for Clinical Microbiome Profiling. *bioRxiv*. 10.1101/2020.08.18.256818
- Bolte, L. A., Klaassen, M. A. Y., Collij, V., Vich Vila, A., Fu, J., van der Meulen, T. A., de Haan, J. J., Versteegen, G. J., Dotinga, A., Zhernakova, A., Wijmenga, C., Weersma, R. K., and Imhann, F. (2021) Patient attitudes towards faecal sampling for gut microbiome studies and clinical care reveal positive engagement and room for improvement. *PLoS One* 16, e0249405
- Lozupone, C. A., Stombaugh, J., Gonzalez, A., Ackermann, G., Wendel, D., Vázquez-Baeza, Y., Jansson, J. K., Gordon, J. I., and Knight, R. (2013) Meta-analyses of studies of the human microbiota. *Genome Res* 23, 1704-1714
- Quince, C., Walker, A. W., Simpson, J. T., Loman, N. J., and Segata, N. (2017) Shotgun metagenomics, from sampling to analysis. *Nat Biotechnol* 35, 833-844
- Wang, Z., Zolnik, C. P., Qiu, Y., Usyk, M., Wang, T., Strickler, H. D., Isasi, C. R., Kaplan, R. C., Kurland, I. J., Qi, Q., and Burk, R. D. (2018) Comparison of Fecal Collection Methods for Microbiome and Metabolomics Studies. *Front Cell Infect Microbiol* 8, 301
- Vogtmann, E., Chen, J., Kibriya, M. G., Chen, Y., Islam, T., Eunes, M., Ahmed, A., Naher, J., Rahman, A., Amir, A., Shi, J., Abnet, C. C., Nelson, H., Knight, R., Chia, N., Ahsan, H., and Sinha, R. (2017) Comparison of Fecal Collection Methods for Microbiota Studies in Bangladesh. *Appl Environ Microbiol* 83
- Wesolowska-Andersen, A., Bahl, M. I., Carvalho, V., Kristiansen, K., Sicheritz-Pontén, T., Gupta, R., and Licht, T. R. (2014) Choice of bacterial DNA extraction method from fecal material influences community structure as evaluated by metagenomic analysis. *Microbiome* 2, 19
- Yuan, S., Cohen, D. B., Ravel, J., Abdo, Z., and Forney, L. J. (2012) Evaluation of methods for the extraction and purification of DNA from the human microbiome. *PLoS One* 7, e33865
- Kennedy, N. A., Walker, A. W., Berry, S. H., Duncan, S. H., Farquarson, F. M., Louis, P., Thomson, J. M., Satsangi, J., Flint, H. J., Parkhill, J., Lees, C. W., and Hold, G. L. (2014) The impact of different DNA extraction kits and laboratories upon the assessment of human gut microbiota composition by 16S rRNA gene sequencing. *PLoS One* 9, e88982
- Tanner, M. A., Goebel, B. M., Dojka, M. A., and Pace, N. R. (1998) Specific ribosomal DNA sequences from diverse environmental settings correlate with experimental contaminants. *Appl Environ Microbiol* 64, 3110-3113
- QIAGEN. (2010) QIAamp® DNA Stool Handbook. For DNA purification from stool samples.
- Schrader, C., Schielke, A., Ellerbroek, L., and Johne, R. (2012) PCR inhibitors - occurrence, properties and removal. *J Appl Microbiol* 113, 1014-1026

© 2022 Mensaje Bioquímico. Todos los derechos reservados. ISSN-0188-137X

Comité Editorial: González Andrade, M.; Hernández Alcántara, G.; Martínez González, J.J.; Meraz Cruz, N.; Ramírez Silva, L.H. y Vilchis Landeros, M.M.

Publicado por el Departamento de Bioquímica de la Facultad de Medicina; UNAM.

Aunque ANCOM y ANCOM-BC tienen características de funcionamiento muy similares en términos de FDR y potencia, ANCOM-BC es computacionalmente más sencillo y rápido de implementar. A diferencia de ANCOM, ANCOM-BC proporciona valores p individuales e intervalos de confianza de la diferencia por pares en la abundancia media para cada taxón. Entre los métodos disponibles actualmente, ANCOM-BC es el único procedimiento que proporciona valores p , e intervalos de confianza válidos. Además, dado que ANCOM-BC se basa en un marco de modelo de regresión, puede ampliarse fácilmente a los ajustes de covariables de medidas repetidas/datos longitudinales (34).

Conclusión

El uso de estas muestras fecales para el estudio de la microbiota intestinal, comprende distintas consideraciones metodológicas a implementar desde su obtención hasta el procesamiento de los datos resultantes, obtenidos mediante metodologías de secuenciación y su análisis. Es por esto que el conocimiento de los retos asociados a la obtención, almacenamiento y procesamiento de muestras fecales y la disponibilidad de técnicas de secuenciación y herramientas bioinformáticas para los análisis de diversidad y taxonomía, puede ayudarnos a mejorar nuestro diseño de experimental al estudiar a la microbiota intestinal.

13. García-Alegría, A. M., Anduro-Corona, I., Pérez-Martínez, C. J., Guadalupe Corella-Madueño, M. A., Rascón-Durán, M. L., and Astiazaran-García, H. (2020) Quantification of DNA through the NanoDrop Spectrophotometer: Methodological Validation Using Standard Reference Material and Sprague Dawley Rat and Human DNA. *Int J Anal Chem* 2020, 8896738
14. Desjardins, P. R., and Conklin, D. S. (2011) Microvolume quantitation of nucleic acids. *Curr Protoc Mol Biol Appendix* 3, 3j
15. Shafikhani, S. (2002) Factors affecting PCR-mediated recombination. *Environ Microbiol* 4, 482-486
16. Herlemann, D. P. R., Labrenz, M., Jürgens, K., Bertilsson, S., Waniek, J. J., and Andersson, A. F. (2011) Transitions in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. *The ISME Journal* 5, 1571-1579
17. Kellogg, D. E., Rybalkin, I., Chen, S., Mukhamedova, N., Vlasik, T., Siebert, P. D., and Chenchik, A. (1994) TaqStart Antibody: "hot start" PCR facilitated by a neutralizing monoclonal antibody directed against Taq DNA polymerase. *Biotechniques* 16, 1134-1137
18. Dietrich, J., Schmitt, P., Zieger, M., Preve, B., Rolland, J. L., Chaabih, H., and Gueguen, Y. (2002) PCR performance of the highly thermostable proof-reading B-type DNA polymerase from *Pyrococcus abyssi*. *FEMS Microbiol Lett* 217, 89-94
19. Gohl, D. M., Vangay, P., Garbe, J., MacLean, A., Hauge, A., Becker, A., Gould, T. J., Clayton, J. B., Johnson, T. J., Hunter, R., Knights, D., and Beckman, K. B. (2016) Systematic improvement of amplicon marker gene methods for increased accuracy in microbiome studies. *Nat Biotechnol* 34, 942-949
20. Bolyen, E., Rideout, J. R., Dillon, M. R., Bokulich, N. A., Abnet, C. C., Al-Ghalith, G. A., Alexander, H., Alm, E. J., Arumugam, M., Asnicar, F., Bai, Y., Bisanz, J. E., Bittinger, K., Brejnrod, A., Brislawn, C. J., Brown, C. T., Callahan, B. J., Caraballo-Rodríguez, A. M., Chase, J., Cope, E. K., Da Silva, R., Diener, C., Dorrestein, P. C., Douglas, G. M., Durall, D. M., Duvallet, C., Edwardson, C. F., Ernst, M., Estaki, M., Fouquier, J., Gauglitz, J. M., Gibbons, S. M., Gibson, D. L., Gonzalez, A., Gorlick, K., Guo, J., Hillmann, B., Holmes, S., Holste, H., Huttenhower, C., Huttley, G. A., Janssen, S., Jarmusch, A. K., Jiang, L., Kaehler, B. D., Kang, K. B., Keefe, C. R., Keim, P., Kelley, S. T., Knights, D., Koester, I., Kosciulek, T., Kreps, J., Langille, M. G. I., Lee, J., Ley, R., Liu, Y. X., Lofthfield, E., Lozupone, C., Maher, M., Marotz, C., Martin, B. D., McDonald, D., McIver, L. J., Melnik, A. V., Metcalf, J. L., Morgan, S. C., Morton, J. T., Naimey, A. T., Navas-Molina, J. A., Nothias, L. F., Orchanian, S. B., Pearson, T., Peoples, S. L., Petras, D., Preuss, M. L., Pruesse, E., Rasmussen, L. B., Rivers, A., Robeson, M. S., 2nd, Rosenthal, P., Segata, N., Shaffer, M., Shiffer, A., Sinha, R., Song, S. J., Spear, J. R., Swafford, A. D., Thompson, L. R., Torres, P. J., Trinh, P., Tripathi, A., Turnbaugh, P. J., Ull-Hasan, S., van der Hooft, J. J. J., Vargas, F., Vázquez-Baeza, Y., Vogtmann, E., von Hippel, M., Walters, W., Wan, Y., Wang, M., Warren, J., Weber, K. C., Williamson, C. H. D., Willis, A. D., Xu, Z. Z., Zaneveld, J. R., Zhang, Y., Zhu, Q., Knight, R., and Caporaso, J. G. (2019) Reproducible, interactive, scalable and extensible microbiome data science using QIIME 2. *Nat Biotechnol* 37, 852-857
21. Vangay, P., Burgin, J., Johnston, A., Beck, K. L., Berrios, D. C., Blumberg, K., Canon, S., Chain, P., Chandonia, J. M., Christianson, D., Costes, S. V., Damerow, J., Duncan, W. D., Dundore-Arias, J. P., Fagnan, K., Galazka, J. M., Gibbons, S. M., Hays, D., Hervey, J., Hu, B., Hurwitz, B. L., Jaiswal, P., Joachimiak, M. P., Kinkel, L., Ladau, J., Martin, S. L., McCue, L. A., Miller, K., Mouncey, N., Mungall, C., Pafilis, E., Reddy, T. B. K., Richardson, L., Roux, S., Schriml, L. M., Shaffer, J. P., Sundaramurthi, J. C., Thompson, L. R., Timme, R. E., Zheng, J., Wood-Charlson, E. M., and Elloe-Fadrosh, E. A. (2021) Microbiome Metadata Standards: Report of the National Microbiome Data Collaborative's Workshop and Follow-On Activities. *mSystems* 6
22. Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., Gonzalez-Beltran, A., Gray, A. J. G., Groth, P., Goble, C., Grethe, J. S., Heringa, J., 't Hoen, P. A. C., Hooft, R., Kuhn, T., Kok, R., Kok, J., Lusher, S. J., Martone, M. E., Mons, A., Packer, A. L., Persson, B., Rocca-Serra, P., Roos, M., van Schaik, R., Sansone, S.-A., Schultes, E., Sengstag, T., Slater, T., Strawn, G., Swertz, M. A., Thompson, M., van der Lei, J., van Mulligen, E., Velterop, J., Waagmeester, A., Wittenburg, P., Wolstencroft, K., Zhao, J., and Mons, B. (2016) The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data* 3, 160018
23. Caporaso, J. G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F. D., Costello, E. K., Fierer, N., Peña, A. G., Goodrich, J. K., Gordon, J. I., Huttley, G. A., Kelley, S. T., Knights, D., Koenig, J. E., Ley, R. E., Lozupone, C. A., McDonald, D., Muegge, B. D., Pirrung, M., Reeder, J., Sevinsky, J. R., Turnbaugh, P. J., Walters, W. A., Widmann, J., Yatsunencko, T., Zaneveld, J., and Knight, R. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* 7, 335-336
24. Callahan, B. J., McMurdie, P. J., and Holmes, S. P. (2017) Exact sequence variants should replace operational taxonomic units in marker-gene data analysis. *The ISME Journal* 11, 2639-2643
25. Sanders, H. L. (1968) Marine Benthic Diversity: A Comparative Study. *The American Naturalist* 102, 243-282
26. Weiss, S., Xu, Z. Z., Peddada, S., Amir, A., Bittinger, K., Gonzalez, A., Lozupone, C., Zaneveld, J. R., Vázquez-Baeza, Y., Birmingham, A., Hyde, E. R., and Knight, R. (2017) Normalization and microbial differential abundance strategies depend upon data characteristics. *Microbiome* 5, 27
27. Lande, R. (1996) Statistics and Partitioning of Species Diversity, and Similarity among Multiple Communities. *Oikos* 76, 5-13
28. Lande, R., DeVries, P. J., and Walla, T. R. (2000) When species accumulation curves intersect: implications for ranking diversity using small samples. *Oikos* 89, 601-605
29. Willis, A. D. (2019) Rarefaction, Alpha Diversity, and Statistics. *Frontiers in Microbiology* 10
30. Anderson, M. J. Permutational Multivariate Analysis of Variance (PERMANOVA). in *Wiley StatsRef: Statistics Reference Online*. pp 1-15
31. CLARKE, K. R. (1993) Non-parametric multivariate analyses of changes in community structure. *Australian Journal of Ecology* 18, 117-143
32. Mandal, S., Van Treuren, W., White, R. A., Eggesbø, M., Knight, R., and Peddada, S. D. (2015) Analysis of composition of microbiomes: a novel method for studying microbial composition. *Microb Ecol Health Dis* 26, 27663
33. Lin, H., and Peddada, S. D. (2020) Analysis of compositions of microbiomes with bias correction. *Nature Communications* 11, 3514
34. Lin, H., and Peddada, S. D. (2020) Analysis of microbial compositions: a review of normalization and differential abundance analysis. *npj Biofilms and Microbiomes* 6, 60



**M. en C. JESSICA FERNANDA
VALDEZ PALOMARES
ORCID: 0000-0002-9238-1191**

La M. en C. Jessica Fernanda Valdez Palomares, originaria del Estado de México, es Licenciada en Nutrición por la Universidad Tecnológica de México graduada con honores. Obtuvo el grado de Maestra en Ciencias en Investigación Clínica Experimental en Salud por la Universidad Nacional Autónoma de México,

con mención honorífica. Ha dedicado su formación al estudio de la microbiota intestinal.

Durante su maestría, estudió la microbiota intestinal de pacientes con síndrome de intestino irritable y su respuesta a la dieta baja en FODMAPs. Ha participado en los congresos nacionales e internacionales FASEB, AMMFEN y Exploring Human Host-Microbiome Interactions in Health and Disease 2020 y en el Consenso Nacional de Cáncer Cervico-uterino 2019 discutiendo el papel inmunomodulador de la microbiota. También ha estudiado la microbiota de esputo de pacientes con Tuberculosis.

Actualmente es candidata a Dra. del Posgrado en Ciencias Biológicas de la UNAM y alumna del Laboratorio de Fisiología de la Nutrición en el Instituto Nacional de Ciencias Médicas y Nutrición Salvador Zubirán, y el Instituto Nacional de Medicina Genómica. Unidad de Vinculación Científica de la Fac. de Medicina UNAM-INMEGEN, donde estudia la microbiota intestinal de niños con desnutrición crónica, y el trasplante de microbiota fecal de donantes humanos a murinos.